REVUE DE STATISTIQUE APPLIQUÉE

P. RIBEREAU

Un modèle linéaire à effets mixtes et applications

Revue de statistique appliquée, tome 54, n° 3 (2006), p. 65-82

<http://www.numdam.org/item?id=RSA_2006__54_3_65_0>

Rev. Statistique Appliquée, 2006, **LIV** (3), 65-82

UN MODÈLE LINÉAIRE À EFFETS MIXTES ET APPLICATIONS

P. RIBEREAU

Université Paris VI, L. S. T. A,Boîte 158, 175, rue du Chevaleret, 75013 Paris, FRANCE,
ribereau@ccr.jussieu.fr

RÉSUMÉ

Nous nous intéressons à un modèle à effets aléatoires défini par :

$$\begin{aligned} X_{i,j} &= m_X + \lambda\left(\mu_{i,j} - m_\mu\right) + \varepsilon_{X;i,j} \\ Y_{i,j} &= m_Y + \left(\mu_{i,j} - m_\mu\right) + \varepsilon_{Y;i,j} \end{aligned}$$

où, pour tout couple (i, j), les variables $\mu_{i,j} - m_\mu$, $\varepsilon_{X;i,j}$ et $\varepsilon_{Y;i,j}$ sont des variables normales centrées et indépendantes. Ce modèle regroupe à la fois les techniques d'analyse de la variance et la structure de modèle linéaire généralisé. Nous nous intéresserons plus particulièrement à l'estimation du paramètre λ et à l'étude des propriétés asymptotiques de cet estimateur. Nous validerons les résultats théoriques sur des données simulées et présenterons une application sur des données betteravières.

Mots-clés : *Modèle linéaire généralisé, analyse de la variance, modèle mixte, données betteravières.*

ABSTRACT

We consider a random effects model defined by :

$$\begin{aligned} X_{i,j} &= m_X + \lambda\left(\mu_{i,j} - m_\mu\right) + \varepsilon_{X;i,j} \\ Y_{i,j} &= m_Y + \left(\mu_{i,j} - m_\mu\right) + \varepsilon_{Y;i,j} \end{aligned}$$

where, for all (i, j), the variables $\mu_{i,j} - m_\mu$, $\varepsilon_{X;i,j}$ and $\varepsilon_{Y;i,j}$ are independent and centered normal random variables. This model combines the ANOVA techniques and the generalized linear model structure. We focus on the estimation of the parameter λ and its asymptotic properties. We illustrate our theoretical results on a simulation study and we present an application on a real data set related to sugar beet.

Keywords : *Generalized linear model, analysis of variance, mixed model, sugar beet data*

1. Introduction et présentation du modèle

L'objet de ce travail est l'étude d'un modèle à effets aléatoires regroupant à la fois la structure de modèle linéaire généralisé (*cf.* McCullagh et Nelder, 1983) et les techniques d'analyse de la variance (*cf.* Dunn et Clark, 1974; Edwards, 1979; Fisher et McDonald, 1978, Scheffé, 1959). Si on note $\tau_{i,j}$ (resp. $\mathbf{t}_{i,j}$) l'observation de la variable aléatoire τ (resp. $\mathbf{t}$) relative à la $j^{ème}$ répétition faite pour le $i^{ème}$ facteur (pour $1 \leqslant i \leqslant I$ et $1 \leqslant j \leqslant n_i$) , le modèle s'écrit :

$$\tau_{i,j} = m_\tau + \lambda\left(\mu_{i,j} - m_\mu\right) + \varepsilon_{\tau,0;i,j} \tag{1}$$

$$t_{i,j} = m_t + \left(\mu_{i,j} - m_\mu\right) + \varepsilon_{t,0;i,j} \tag{2}$$

où, pour tout couple (i,j) donné, les variables aléatoires $\varepsilon_{\tau,0;i,j} \equiv \mathcal{N}\left(0, \sigma^2_{\tau,0}\right)$, $\varepsilon_{t,0;i,j} \equiv \mathcal{N}\left(0, \sigma^2_{t,0}\right)$ et $\mu_{i,j} \equiv \mathcal{N}\left(m_\mu, \sigma^2_\mu\right)$ sont supposés normales et indépendantes, λ, m_τ, $m_{\mathbf{t}}$, m_μ, $\sigma^2_{\tau,0}$ et $\sigma^2_{\mathbf{t},0}$ étant des paramètres à estimer.

Nous représentons la variation de $\mu_{i,j}$ en fonction de i et j par la décomposition suivante :

$$\mu_{i,j} = \mu_{s;i} + \mu_{g;i,j} \tag{3}$$

où les facteurs additifs de la décomposition ont la signification suivante :

$$\mu_{s;i} \equiv \mathcal{N}\left(0, \sigma^2_{\mu,s}\right) = \text{variation relative du facteur } i \text{ par rapport à la moyenne générale} \tag{4}$$

$$\mu_{g;i,j} \equiv \mathcal{N}\left(m_\mu, \sigma^2_{\mu,g}\right) = \text{variation relative de la } j^{ème} \text{ repétition par rapport à la moyenne du } i^{ème} \text{ facteur.} \tag{5}$$

Les deux variables sont supposées normales et indépendantes. On a aussi la décomposition suivante :

$$\varepsilon_{t,0;i,j} = \varepsilon_{t,s;i} + \varepsilon_{t,g;i,j} \tag{6}$$

où les variables $\varepsilon_{t,s;i} \equiv \mathcal{N}\left(0, \sigma^2_{t,s}\right)$ et $\varepsilon_{t,g;i,j} \equiv \mathcal{N}\left(0, \sigma^2_{t,g}\right)$ sont supposées indépendantes.

On observe que :

$$\text{Var}(\tau_{i,j}) = \lambda^2\sigma^2_\mu + \sigma^2_{\tau,0} \tag{7}$$

$$\text{Var}(\mathbf{t}_{i,j}) = \sigma^2_\mu + \sigma^2_{\mathbf{t},0} \tag{8}$$

et

$$\text{Cov}(\tau_{i,j}, \mathbf{t}_{i,j}) = \lambda\,\sigma^2_\mu \tag{9}$$

On remarque également que $(\tau_{i,j}, \mathbf{t}_{i,j})$ et $(\tau_{r,s}, \mathbf{t}_{r,s})$ sont indépendants pour $i \neq r$, c'est-à-dire que les mesures sont indépendantes d'un facteur à l'autre. Toutefois, si $i = r$ et $j \neq s$, on constate que :

$$\mathrm{Cov}(\tau_{i,j}, \tau_{i,s}) = \lambda^2 \sigma^2_{\mu,s} \tag{10}$$

$$\mathrm{Cov}(\mathbf{t}_{i,j}, \mathbf{t}_{i,s}) = \sigma^2_{\mu,s} \tag{11}$$

et

$$\mathrm{Cov}(\tau_{i,j}, \mathbf{t}_{i,s}) = \lambda \sigma^2_{\mu,s} \tag{12}$$

Enfin, pour rendre le modèle identifiable, il nous faut supposer que les quantités m_μ, m_τ et $m_\mathbf{t}$ sont égales. Nous allons nous intéresser dans la section suivante à l'estimation des différents paramètres du modèle.

Remarque. – Si on pose $m_\mu = 0$ ce qui implique que $\mu_{i,j}$ est centrée, on obtient le même modèle en le rendant identifiable. On peut alors estimer $m_\mathbf{t}$ et m_τ par $\overline{t}_{\bullet,\bullet}$ et $\overline{\tau}_{\bullet,\bullet}$ respectivement (définis en (13) et (14)). Ce choix n'a pas été fait ici à la demande des industriels pour l'application betteravière motivant ce travail.

Ces travaux sont motivés par une application pratique concernant la détermination automatique du taux de collet d'une betterave qui sera présentée en Section 4.2. Nous nous intéresserons dans un premier temps à l'estimation des différents paramètres avant de focaliser notre attention sur le paramètre λ. Nous établirons les propriétés asymptotiques de l'estimateur de λ sous certaines conditions avant de les valider sur des simulations dans la Section 4.1.

2. Estimation des paramètres

Soit $n = \sum_{i=1}^{I} n_i$. Posons

$$\overline{\tau}_{i,\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i} \tau_{i,j}, \quad \overline{\tau}_{\bullet,\bullet} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i} \tau_{i,j} = \frac{1}{n}\sum_{i=1}^{I} n_i \overline{\tau}_{i,\bullet} \tag{13}$$

$$\overline{t}_{i,\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i} t_{i,j}, \quad \overline{t}_{\bullet,\bullet} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i} t_{i,j} = \frac{1}{n}\sum_{i=1}^{I} n_i \overline{t}_{i,\bullet} \tag{14}$$

Posons également

$$\overline{\varepsilon}_{\tau,0;i,\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i} \varepsilon_{\tau,0;i,j} \equiv \mathcal{N}\left(0, \frac{1}{n_i}\sigma^2_{\tau,0}\right) \tag{15}$$

$$\overline{\varepsilon}_{\tau,0;\bullet,\bullet} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\varepsilon_{\tau,0;i,j} \equiv \mathcal{N}\left(0, \frac{1}{n}\sigma^2_{\tau,0}\right) \tag{16}$$

$$\overline{\varepsilon}_{t,0;i,\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i}\varepsilon_{t,0;i,j} \equiv \mathcal{N}\left(0, \sigma^2_{t,s} + \frac{1}{n_i}\sigma^2_{t,g}\right) \tag{17}$$

$$\overline{\varepsilon}_{t,0;\bullet,\bullet} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\varepsilon_{t,0;i,j} \equiv \mathcal{N}\left(0, \frac{1}{n^2}\left\{\sum_{i=1}^{I} n_i^2\right\}\sigma^2_{t,s} + \frac{1}{n}\sigma^2_{t,g}\right) \tag{18}$$

et enfin,

$$\overline{\mu}_{g;i,\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i}\mu_{g;i,j} \equiv \mathcal{N}\left(m_\mu, \frac{1}{n_i}\sigma^2_{\mu,g}\right) \tag{19}$$

$$\overline{\mu}_{g;\bullet,\bullet} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\mu_{g;i,j} \equiv \mathcal{N}\left(m_\mu, \frac{1}{n}\sigma^2_{\mu,g}\right) \tag{20}$$

$$\widetilde{\mu}_{s;\bullet} = \frac{1}{n}\sum_{i=1}^{I} n_i\mu_{s;i} \equiv \mathcal{N}\left(0, \frac{1}{n^2}\left\{\sum_{i=1}^{I} n_i^2\right\}\sigma^2_{\mu,s}\right) \tag{21}$$

$$\overline{\mu}_{i,\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i}\mu_{i,j} = \mu_{s;i} + \overline{\mu}_{g:i,\bullet} \equiv \mathcal{N}\left(m_\mu, \sigma^2_{\mu,s} + \frac{1}{n_i}\sigma^2_{\mu,g}\right) \tag{22}$$

$$\begin{aligned}\overline{\mu}_{\bullet,\bullet} &= \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\mu_{i,j} = \widetilde{\mu}_{s;\bullet} + \overline{\mu}_{g;\bullet,\bullet} \\ &\equiv \mathcal{N}\left(m_\mu, \frac{1}{n^2}\left\{\sum_{i=1}^{I} n_i^2\right\}\sigma^2_{\mu,s} + \frac{1}{n}\sigma^2_{\mu,g}\right)\end{aligned} \tag{23}$$

On constate que :

$$\tau_{i,j} - \overline{\tau}_{i,\bullet} = \lambda\left(\mu_{g;i,j} - \overline{\mu}_{g;i,\bullet}\right) + \varepsilon_{\tau,0;i,j} - \overline{\varepsilon}_{\tau,0;i,\bullet} \tag{24}$$

$$t_{i,j} - \overline{t}_{i,\bullet} = \mu_{g;i,j} - \overline{\mu}_{g;i,\bullet} + \varepsilon_{t,0;i,j} - \overline{\varepsilon}_{t,0;i,\bullet} \tag{25}$$

$$\overline{\tau}_{i,\bullet} - \overline{\tau}_{\bullet,\bullet} = \lambda\left(\mu_{s;i} - \widetilde{\mu}_{s;\bullet} + \overline{\mu}_{g;i,\bullet} - \overline{\mu}_{g;\bullet,\bullet}\right) + \overline{\varepsilon}_{\tau,0;i,\bullet} - \overline{\varepsilon}_{\tau,0;\bullet,\bullet} \tag{26}$$

$$\overline{t}_{i,\bullet} - \overline{t}_{\bullet,\bullet} = \mu_{s;i} - \widetilde{\mu}_{s;\bullet} + \overline{\mu}_{g;i,\bullet} - \overline{\mu}_{g;\bullet,\bullet} + \overline{\varepsilon}_{t,0;i,\bullet} - \overline{\varepsilon}_{t,0;\bullet,\bullet} \tag{27}$$

De plus, nous avons les relations suivantes :

$$\tau_{i,j} - \overline{\tau}_{i,\bullet} = (\tau_{i,j} - \lambda\mu_{s;i}) - (\overline{\tau}_{i,\bullet} - \lambda\mu_{s;i}),$$

$$t_{i,j} - \overline{t}_{i,\bullet} = (t_{i,j} - \mu_{s;i} - \varepsilon_{t;s;i}) - \left(\overline{t}_{i,\bullet} - \mu_{s;i} - \varepsilon_{t;s;i}\right).$$

Comme

$$\begin{aligned} \mathrm{Var}\,(\tau_{i,j} - \lambda\mu_{s;i}) &= \mathrm{Var}\,(\lambda\mu_{g;i,j} + \varepsilon_{\tau,0;i,j}) \\ &= \lambda^2\sigma^2_{\mu,g} + \sigma^2_{\tau,0} \qquad (28) \\ \mathrm{Var}\,(t_{i,j} - \mu_{s;i} - \varepsilon_{t;s;i}) &= \mathrm{Var}\,(\mu_{g;i,j} + \varepsilon_{t,g;i,j}) \\ &= \sigma^2_{\mu,g} + \sigma^2_{t,g} \qquad (29) \\ \mathrm{Cov}\,(\tau_{i,j} - \lambda\mu_{s;i}, t_{i,j} - \mu_{s;i} - \varepsilon_{t;s;i}) &= \lambda \mathrm{Var}\,(\mu_{g;i,j}), \\ &= \lambda\sigma^2_{\mu,g} \qquad (30) \end{aligned}$$

on en déduit donc que :

$$\frac{1}{\lambda^2\sigma^2_{\mu,g} + \sigma^2_{\tau,0}} \sum_{j=1}^{n_i} (\tau_{i,j} - \overline{\tau}_{i,\bullet})^2 \equiv \chi^2_{n_i-1} \qquad (31)$$

$$\frac{1}{\sigma^2_{\mu,g} + \sigma^2_{t,g}} \sum_{j=1}^{n_i} \left(t_{i,j} - \overline{t}_{i,\bullet}\right)^2 \equiv \chi^2_{n_i-1} \qquad (32)$$

Plus généralement, on constate, en posant :

$$R_{11} = \frac{1}{n-I} \sum_{i=1}^{I} \sum_{j=1}^{n_i} (\tau_{i,j} - \overline{\tau}_{i,\bullet})^2 \qquad (33)$$

$$R_{22} = \frac{1}{n-I} \sum_{i=1}^{I} \sum_{j=1}^{n_i} \left(t_{i,j} - \overline{t}_{i,\bullet}\right)^2 \qquad (34)$$

et

$$R_{12} = \frac{1}{n-I} \sum_{i=1}^{I} \sum_{j=1}^{n_i} (\tau_{i,j} - \overline{\tau}_{i,\bullet}) \left(t_{i,j} - \overline{t}_{i,\bullet}\right) \qquad (35)$$

que la matrice

$$(n-I) \begin{bmatrix} R_{11} & R_{12} \\ R_{12} & R_{22} \end{bmatrix} \equiv W_2\,(n-I, \Sigma_1) \qquad (36)$$

suit une loi de Wishart à $n-I$ degrés de libertés, associée à la matrice

$$\Sigma_1 = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \lambda^2\sigma^2_{\mu,g} + \sigma^2_{\tau,0} & \lambda\sigma^2_{\mu,g} \\ \lambda\sigma^2_{\mu,g} & \sigma^2_{\mu,g} + \sigma^2_{t,g} \end{bmatrix} \tag{37}$$

Introduisons maintenant les statistiques

$$\begin{aligned} T_{11} &= \frac{1}{I-1}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(\overline{\tau}_{i,\bullet} - \overline{\tau}_{\bullet,\bullet}\right)^2 \\ &= \frac{1}{I-1}\sum_{i=1}^{I} n_i\left(\overline{\tau}_{i,\bullet} - \overline{\tau}_{\bullet,\bullet}\right)^2 \end{aligned} \tag{38}$$

$$\begin{aligned} T_{22} &= \frac{1}{I-1}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(\overline{t}_{i,\bullet} - \overline{t}_{\bullet,\bullet}\right)^2 \\ &= \frac{1}{I-1}\sum_{i=1}^{I} n_i\left(\overline{t}_{i,\bullet} - \overline{t}_{\bullet,\bullet}\right)^2 \end{aligned} \tag{39}$$

et

$$\begin{aligned} T_{12} &= \frac{1}{I-1}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(\overline{\tau}_{i,\bullet} - \overline{\tau}_{\bullet,\bullet}\right)\left(\overline{t}_{i,\bullet} - \overline{t}_{\bullet,\bullet}\right) \\ &= \frac{1}{I-1}\sum_{i=1}^{I} n_i\left(\overline{\tau}_{i,\bullet} - \overline{\tau}_{\bullet,\bullet}\right)\left(\overline{t}_{i,\bullet} - \overline{t}_{\bullet,\bullet}\right). \end{aligned} \tag{40}$$

Dans le cas particulier où $n_j = n_1$ pour tout $1 \leqslant j \leqslant I$, on a :

$$\frac{1}{\lambda^2\left\{\sigma^2_{\mu,s} + \frac{1}{n_1}\sigma^2_{\mu,g}\right\} + \frac{1}{n_1}\sigma^2_{\tau,0}} \sum_{i=1}^{I}\left(\overline{\tau}_{i,\bullet} - \overline{\tau}_{\bullet,\bullet}\right)^2 \equiv \chi^2_{I-1} \tag{41}$$

$$\frac{1}{\sigma^2_{\mu,s} + \frac{1}{n_1}\sigma^2_{\mu,g} + \sigma^2_{t,s} + \frac{1}{n_1}\sigma^2_{t,g}} \sum_{i=1}^{I}\left(\overline{t}_{i,\bullet} - \overline{t}_{\bullet,\bullet}\right)^2 \equiv \chi^2_{I-1} \tag{42}$$

Dans le cas général, commençons par constater que

$$\begin{aligned}\overline{\tau}_{i,\bullet} - m_\tau &= \lambda\left(\mu_{s;i} + \overline{\mu}_{g;i,\bullet} - m_\mu\right) + \overline{\varepsilon}_{\tau,0;i,\bullet} \\ &\equiv \mathcal{N}\left(0, \lambda^2\left\{\sigma^2_{\mu,s} + \frac{1}{n_i}\sigma^2_{\mu,g}\right\} + \frac{1}{n_i}\sigma^2_{\tau,0}\right) \qquad (43)\end{aligned}$$

$$\begin{aligned}\overline{t}_{i,\bullet} - m_t &= \mu_{s;i} + \overline{\mu}_{g;i,\bullet} - m_\mu + \overline{\varepsilon}_{t,0;i,\bullet} \\ &\equiv \mathcal{N}\left(0, \left\{\sigma^2_{\mu,s} + \frac{1}{n_i}\sigma^2_{\mu,g}\right\} + \sigma^2_{t,s} + \frac{1}{n_i}\sigma^2_{t,g}\right) \qquad (44)\end{aligned}$$

$$\begin{aligned}\overline{\tau}_{\bullet,\bullet} - m_\tau &= \lambda\left(\widetilde{\mu}_{s;\bullet} + \overline{\mu}_{g;\bullet,\bullet} - m_\mu\right) + \overline{\varepsilon}_{\tau,0;\bullet,\bullet} \\ &\equiv \mathcal{N}\left(0, \frac{1}{n^2}\left\{\sum_{i=1}^{I} n_i^2\right\}\lambda^2\sigma^2_{\mu,s} + \frac{1}{n}\left\{\lambda^2\sigma^2_{\mu,g} + \sigma^2_{\tau,0}\right\}\right) \qquad (45)\end{aligned}$$

$$\begin{aligned}\overline{t}_{\bullet,\bullet} - m_t &= \widetilde{\mu}_{s;\bullet} + \overline{\mu}_{g;\bullet,\bullet} - m_\mu + \overline{\varepsilon}_{t,0;\bullet,\bullet} \\ &\equiv \mathcal{N}\left(0, \frac{1}{n^2}\left\{\sum_{i=1}^{I} n_i^2\right\}\left(\sigma^2_{\mu,s} + \sigma^2_{t,s}\right) + \frac{1}{n}\left\{\sigma^2_{\mu,g} + \sigma^2_{t,g}\right\}\right) \qquad (46)\end{aligned}$$

On peut aussi remarquer que

$$\mathbb{E}\left(\left(\overline{\tau}_{i,\bullet} - m_\tau\right)\left(\overline{\tau}_{\bullet,\bullet} - m_\tau\right)\right) = \frac{n_i}{n}\lambda^2\sigma^2_{\mu,s} + \frac{1}{n}\left\{\lambda^2\sigma^2_{\mu,g} + \sigma^2_{\tau,0}\right\} \qquad (47)$$

$$\mathbb{E}\left(\left(\overline{t}_{i,\bullet} - m_t\right)\left(\overline{t}_{\bullet,\bullet} - m_t\right)\right) = \frac{n_i}{n}\left\{\sigma^2_{\mu,s} + \sigma^2_{t,s}\right\} + \frac{1}{n}\left\{\sigma^2_{\mu,g} + \sigma^2_{t,g}\right\} \qquad (48)$$

$$\mathbb{E}\left(\left(\overline{\tau}_{i,\bullet} - m_\tau\right)\left(\overline{t}_{i,\bullet} - m_t\right)\right) = \lambda\sigma^2_{\mu,s} + \frac{1}{n_i}\lambda\sigma^2_{\mu,g} \qquad (49)$$

$$\mathbb{E}\left(\left(\overline{\tau}_{i,\bullet} - m_\tau\right)\left(\overline{t}_{\bullet,\bullet} - m_t\right)\right) = \frac{n_i}{n}\lambda\sigma^2_{\mu,s} + \frac{1}{n}\lambda\sigma^2_{\mu,g} \qquad (50)$$

$$\mathbb{E}\left(\left(\overline{t}_{i,\bullet} - m_t\right)\left(\overline{\tau}_{\bullet,\bullet} - m_\tau\right)\right) = \frac{n_i}{n}\lambda\sigma^2_{\mu,s} + \frac{1}{n}\lambda\sigma^2_{\mu,g} \qquad (51)$$

$$\mathbb{E}\left(\left(\overline{\tau}_{\bullet,\bullet} - m_\tau\right)\left(\overline{t}_{\bullet,\bullet} - m_t\right)\right) = \frac{1}{n^2}\left\{\sum_{i=1}^{I} n_i^2\right\}\lambda\sigma^2_{\mu,s} + \frac{1}{n}\lambda\sigma^2_{\mu,g} \qquad (52)$$

De ces résultats, on peut facilement déduire que

$$\begin{aligned}
\mathbb{E}\left((I-1)\,T_{11}\right) &= \mathbb{E}\left(\sum_{i=1}^{I} n_i\left(\overline{\tau}_{i,\bullet}-\overline{\tau}_{\bullet,\bullet}\right)^2\right)\\
&= \sum_{i=1}^{I} n_i \mathbb{E}\left\{\left(\overline{\tau}_{i,\bullet}-m_\tau\right)^2+\left(\overline{\tau}_{\bullet,\bullet}-m_\tau\right)^2-2\left(\overline{\tau}_{i,\bullet}-m_\tau\right)\left(\overline{\tau}_{\bullet,\bullet}-m_\tau\right)\right\}\\
&= \left\{n-\frac{1}{n}\sum_{i=1}^{I} n_i^2\right\}\lambda^2\sigma_{\mu,s}^2+\{I-1\}\left\{\lambda^2\sigma_{\mu,g}^2+\sigma_{\tau,0}^2\right\}\\
&= \left\{n-\frac{1}{n}\sum_{i=1}^{I} n_i^2\right\}\lambda^2\sigma_{\mu,s}^2+\{I-1\}\,\mathbb{E}\left(R_{11}\right) \qquad (53)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left((I-1)\,T_{22}\right) &= \mathbb{E}\left(\sum_{i=1}^{I} n_i\left(\overline{t}_{i,\bullet}-\overline{t}_{\bullet,\bullet}\right)^2\right)\\
&= \sum_{i=1}^{I} n_i \mathbb{E}\left\{\left(\overline{t}_{i,\bullet}-m_t\right)^2+\left(\overline{t}_{\bullet,\bullet}-m_t\right)^2-2\left(\overline{t}_{i,\bullet}-m_t\right)\left(\overline{t}_{\bullet,\bullet}-m_t\right)\right\}\\
&= \left\{n-\frac{1}{n}\sum_{i=1}^{I} n_i^2\right\}\left(\sigma_{\mu,s}^2+\sigma_{t,s}^2\right)+\{I-1\}\left\{\sigma_{\mu,g}^2+\sigma_{t,g}^2\right\}\\
&= \left\{n-\frac{1}{n}\sum_{i=1}^{I} n_i^2\right\}\left(\sigma_{\mu,s}^2+\sigma_{t,s}^2\right)+\{I-1\}\,\mathbb{E}\left(R_{22}\right) \qquad (54)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left((I-1)\,T_{12}\right) &= \mathbb{E}\left(\sum_{i=1}^{I} n_i\left(\overline{\tau}_{i,\bullet}-\overline{\tau}_{\bullet,\bullet}\right)\left(\overline{t}_{i,\bullet}-\overline{t}_{\bullet,\bullet}\right)\right)\\
&= \sum_{i=1}^{I} n_i \mathbb{E}\left\{\left(\overline{\tau}_{i,\bullet}-m_\tau\right)\left(\overline{t}_{i,\bullet}-m_t\right)-\left(\overline{\tau}_{i,\bullet}-m_\tau\right)\left(\overline{t}_{\bullet,\bullet}-m_t\right)\right.\\
&\qquad\left.+\left(\overline{\tau}_{\bullet,\bullet}-m_\tau\right)\left(\overline{t}_{\bullet,\bullet}-m_t\right)-\left(\overline{\tau}_{\bullet,\bullet}-m_\tau\right)\left(\overline{t}_{i,\bullet}-m_t\right)\right\}\\
&= \left\{n-\frac{1}{n}\sum_{i=1}^{I} n_i^2\right\}\lambda\sigma_{\mu,s}^2+\{I-1\}\,\lambda\sigma_{\mu,g}^2\\
&= \left\{n-\frac{1}{n}\sum_{i=1}^{I} n_i^2\right\}\lambda\sigma_{\mu,s}^2+\{I-1\}\,\mathbb{E}\left(R_{12}\right) \qquad (55)
\end{aligned}$$

On constate donc que

$$\mathbb{E}\left(T_{11} - R_{11}\right) = \frac{1}{I-1}\left\{n - \frac{1}{n}\sum_{i=1}^{I} n_i^2\right\}\lambda^2\sigma_{\mu,s}^2 \tag{56}$$

$$\mathbb{E}\left(T_{22} - R_{22}\right) = \frac{1}{I-1}\left\{n - \frac{1}{n}\sum_{i=1}^{I} n_i^2\right\}\left\{\sigma_{\mu,s}^2 + \sigma_{t,s}^2\right\} \tag{57}$$

$$\mathbb{E}\left(T_{12} - R_{12}\right) = \frac{1}{I-1}\left\{n - \frac{1}{n}\sum_{i=1}^{I} n_i^2\right\}\lambda\sigma_{\mu,s}^2 \tag{58}$$

Posons

$$L = \frac{1}{I-1}\left\{n - \frac{1}{n}\sum_{i=1}^{I} n_i^2\right\} \tag{59}$$

Notons que cette quantité se réduit à n_1 si tous les n_i sont égaux ($n_i = n_1$ quelque soit i). A partir des formules (36), (37), (56) à (58), on obtient les estimateurs naturels de λ, $\sigma_{\mu,s}^2$, $\sigma_{t,s}^2$, $\sigma_{\mu,g}^2$, $\sigma_{t,g}^2$ et de $\sigma_{\tau,0}^2$ avec les formules suivantes

$$\widehat{\lambda} = \frac{T_{11} - R_{11}}{T_{12} - R_{12}} \tag{60}$$

$$\widehat{\sigma}_{\mu,s}^2 = \frac{1}{L} \times \frac{\left(T_{12} - R_{12}\right)^2}{T_{11} - R_{11}} \tag{61}$$

$$\widehat{\sigma}_{t,s}^2 = \frac{1}{L} \times \left\{T_{22} - R_{22} - \frac{\left(T_{12} - R_{12}\right)^2}{T_{11} - R_{11}}\right\} \tag{62}$$

$$\widehat{\sigma}_{\mu,g}^2 = R_{12} \times \left\{\frac{T_{12} - R_{12}}{T_{11} - R_{11}}\right\} \tag{63}$$

$$\widehat{\sigma}_{t,g}^2 = R_{22} - R_{12} \times \left\{\frac{T_{12} - R_{12}}{T_{11} - R_{11}}\right\} \tag{64}$$

$$\widehat{\sigma}_{\tau,0}^2 = R_{11} - R_{12} \times \left\{\frac{T_{11} - R_{11}}{T_{12} - R_{12}}\right\} \tag{65}$$

Enfin, l'estimateur de m_μ est obtenu par :

$$\widehat{m}_\mu = \frac{\overline{t}_{\bullet,\bullet} + \overline{\tau}_{\bullet,\bullet}}{2} \tag{66}$$

Le paramètre le plus important pour nous est le paramètre λ qui caractérise la dépendance entre les deux variables. Nous allons donc nous intéresser, dans la section suivante, à la loi limite de son estimateur $\widehat{\lambda}$. Toutefois un travail similaire pourrait être fait pour les autres paramètres (variances) du modèle.

Remarque. – Dans l'application, les variables $\mathbf{t}$ et τ sont deux mesures différentes du taux de collet d'un lot de betteraves. Dans ce cas là, le paramètre λ peut être interprété comme un paramètre de conformité entre les deux mesures. Ainsi, plus λ sera proche de 1, plus les mesures pourront être jugées conformes entre elles.

3. Loi limite de $\widehat{\lambda}$

Le paramètre λ joue un rôle crucial dans la mesure où il caractérise la dépendance entre les deux variables. Nous allons donc nous intéresser à la loi limite de son estimateur $\widehat{\lambda}$. Pour cela, nous supposerons que tous les n_i sont égaux. Dans ce cas là, les variables aléatoires $\overline{t}_{i,\bullet}$ et $\overline{\tau}_{i,\bullet}$ sont des variables indépendantes et identiquement distribuées. On a alors, compte tenu des formules (43) et (44)

$$\overline{\tau}_{i,\bullet} \equiv \mathcal{N}\left(m_\tau, \lambda^2\left(\sigma^2_{\mu,s} + \frac{1}{n_1}\sigma^2_{\mu,g}\right) + \frac{1}{n_1}\sigma^2_{\tau,0}\right) \tag{67}$$

$$\overline{t}_{i,\bullet} \equiv \mathcal{N}\left(m_t, \sigma^2_{\mu,s} + \frac{1}{n_1}\sigma^2_{\mu,g} + \frac{1}{n_1}\sigma^2_{t,g} + \sigma^2_{t,s}\right) \tag{68}$$

et donc

$$\frac{1}{\lambda^2(\sigma^2_{\mu,s} + \frac{1}{n_1}\sigma^2_{\mu,g}) + \frac{1}{n_1}\sigma^2_{\tau,0}} \sum_{i=1}^{I} (\overline{\tau}_{i,\bullet} - \overline{\tau}_{\bullet,\bullet})^2 \equiv \chi^2_{I-1} \tag{69}$$

$$\frac{1}{\sigma^2_{\mu,s} + \frac{1}{n_1}\sigma^2_{\mu,g} + \frac{1}{n_1}\sigma^2_{t,g} + \sigma^2_{t,s}} \sum_{i=1}^{I} \left(\overline{t}_{i,\bullet} - \overline{t}_{\bullet,\bullet}\right)^2 \equiv \chi^2_{I-1} \tag{70}$$

On peut remarquer aussi que

$$\mathrm{Cov}(\overline{t}_{i,\bullet}, \overline{\tau}_{i,\bullet}) = \lambda\left(\sigma^2_{\mu,s} + \frac{1}{n_1}\sigma^2_{\mu,g}\right) \tag{71}$$

On en déduit donc que la matrice :

$$(I-1)\begin{bmatrix} T_{11} & T_{12} \\ T_{12} & T_{22} \end{bmatrix} \equiv W_2\left(I-1, \Sigma_2\right) \tag{72}$$

suit une loi de Wishart à $I-1$ degrés de liberté, associée à la matrice

$$\Sigma_2=\begin{bmatrix}\sigma'_{11} & \sigma'_{12}\\ \sigma'_{12} & \sigma'_{22}\end{bmatrix}=\begin{bmatrix}\lambda^2 n_1\left(\sigma^2_{\mu,s}+\frac{1}{n_1}\sigma^2_{\mu,g}\right)+\sigma^2_{\tau,0} & \lambda n_1\left(\sigma^2_{\mu,s}+\frac{1}{n_1}\sigma^2_{\mu,g}\right)\\ \lambda n_1\left(\sigma^2_{\mu,s}+\frac{1}{n_1}\sigma^2_{\mu,g}\right) & n_1\sigma^2_{\mu,s}+\sigma^2_{\mu,g}+\sigma^2_{t,g}+n_1\sigma^2_{t,s}\end{bmatrix} \tag{73}$$

On rappelle aussi que

$$(n-I)\begin{bmatrix}R_{11} & R_{12}\\ R_{12} & R_{22}\end{bmatrix}\equiv W_2\left(n-I,\Sigma_1\right) \tag{74}$$

suit une loi de Wishart à $n-I$ degrés de libertés.

De plus

$$\hat{\lambda}=\frac{T_{11}-R_{11}}{T_{12}-R_{12}} \tag{75}$$

Les méthodes d'estimation des T_{ij} et R_{ij} sont des méthodes des moments et, dans ce cas là, coïncident avec les méthodes d'estimation du maximum de vraisemblance. Tous les moments empiriques sont asymptotiquement normaux (*cf.* Kendall *et al.*, 1987). Donc les T_{ij} et R_{ij} sont tous asymptotiquement normaux. Ceci nous conduit au résultat suivant :

THÉORÈME 1. – *Considérons le modèle défini en Section 1 et supposons les n_i tous identiques à n_1. Alors, on a*

$$\sqrt{I}\left(\widehat{\lambda}-\lambda\right)\xrightarrow{d}\mathcal{N}(0,\sigma^2)\quad \textit{quand } I\to\infty \tag{76}$$

où $\xrightarrow{d}$ correspond à la convergence en loi, avec

$$\sigma^2=\frac{1}{n_1^2\lambda^2\sigma^4_{\mu,s}}\left\{\lambda^2\left(A+B\right)+\left(C+D\right)-2\,\lambda(E+F)\right\} \tag{77}$$

où

$$\begin{aligned}A=&\left[\lambda^2 n_1\left(\sigma^2_{\mu,s}+\frac{1}{n_1}\sigma^2_{\mu,g}\right)+\sigma^2_{\tau,0}\right]\left[n_1\sigma^2_{\mu,s}+\sigma^2_{\mu,g}+\sigma^2_{t,g}+n_1\sigma^2_{t,s}\right]\\ &+\lambda^2 n_1^2\left[\sigma^2_{\mu,s}+\frac{1}{n_1}\sigma^2_{\mu,g}\right]^2\end{aligned} \tag{78}$$

$$B=\left\{\left[\lambda^2\sigma^2_{\mu,g}+\sigma^2_{\tau,0}\right]\left[\sigma^2_{\mu,g}+\sigma^2_{t,g}\right]+\lambda^2\sigma^4_{\mu,g}\right\}\Big/(n_1-1) \tag{79}$$

$$C = 2\left[\lambda^2 n_1\left(\sigma^2_{\mu,s} + \frac{1}{n_1}\sigma^2_{\mu,g}\right) + \sigma^2_{\tau,0}\right]^2 \qquad (80)$$

$$D = 2\left(\lambda^2\sigma^2_{\mu,g} + \sigma^2_{\tau,0}\right)^2 \Big/ (n_1 - 1) \qquad (81)$$

$$E = 2\left\{\lambda n_1\left[\lambda^2 n_1\left(\sigma^2_{\mu,s} + \frac{1}{n_1}\sigma^2_{\mu,g}\right) + \sigma^2_{\tau,0}\right]\left[\sigma^2_{\mu,s} + \frac{1}{n_1}\sigma^2_{\mu,g}\right]\right\} \qquad (82)$$

$$F = 2\left\{\lambda\sigma^2_{\mu,g}\left[\lambda^2\sigma^2_{\mu,g} + \sigma^2_{\tau,0}\right]\right\} \Big/ (n_1 - 1) \qquad (83)$$

La preuve de ce résultat sera donnée en Section 5.

Afin d'apprécier la qualité du modèle, nous allons effectuer, dans la section suivante, des simulations et analyser les résultats obtenus dans le cas des données réelles betteravières.

4. Simulations et résultats

4.1. Données simulées

Nous avons simulé avec le logiciel R des échantillons suivant les formules définies dans la Section 1 en rentrant les valeurs de m_μ, $\sigma^2_{\mu,s}$, $\sigma^2_{\mu,g}$, $\sigma^2_{\tau,0}$, $\sigma^2_{t,g}$ et $\sigma^2_{t,s}$. Dans les Tableaux 1-3, on a, pour chaque jeu de ces 6 paramètres, simulé 1000 échantillons avec $I = 100$ et $n_1 = 30$ pour $\lambda = 1, 0.7, 0.5$ et 0.3. À chaque fois, on calcule la moyenne des estimations et on compare l'écart-type empirique et l'écart-type théorique donné par les équations (77) à (83).

TABLEAU 1
1 000 *répétitions d'échantillons avec* $I = 100$ *et* $n_1 = 30$ *pour* $m_\mu = 10$, $\sigma^2_{\mu,s} = 1$, $\sigma^2_{\mu,g} = 1$, $\sigma^2_{\tau,0} = 1$, $\sigma^2_{t,s} = 1$ *et* $\sigma^2_{t,g} = 1$

λ	Moyenne des $\hat{\lambda}_i$	Écart-type empirique	Écart-type théorique
1	1.005696	0.1104061	0.1068603
0.7	0.7036715	0.07918313	0.07716930
0.5	0.5031209	0.06324146	0.05848543
0.3	0.3023484	0.04738571	0.04322715

TABLEAU 2
1 000 *répétitions d'échantillons avec* $I = 100$ *et* $n_1 = 30$ *pour* $m_\mu = 10$, $\sigma^2_{\mu,s} = 1.5$, $\sigma^2_{\mu,g} = 0.25$, $\sigma^2_{\tau,0} = 1.8$, $\sigma^2_{t,s} = 1$ *et* $\sigma^2_{t,g} = 2$

λ	Moyenne des $\hat{\lambda}_i$	Écart-type empirique	Écart-type théorique
1	1.014882	0.09108891	0.08698713
0.7	0.7005774	0.06200009	0.06258740
0.5	0.5053948	0.04975500	0.04706345
0.3	0.3016827	0.03396491	0.03376581

TABLEAU 3
1 000 *répétitions d'échantillons avec* $I = 100$ *et* $n_1 = 30$ *pour* $m_\mu = 20$, $\sigma^2_{\mu,s} = 2$, $\sigma^2_{\mu,g} = 0.5$, $\sigma^2_{\tau,0} = 1.5$, $\sigma^2_{t,s} = 0.5$ *et* $\sigma^2_{t,g} = 2$

λ	Moyenne des $\hat{\lambda}_i$	Écart-type empirique	Écart-type théorique
1	1.00127	0.05641955	0.0566791
0.7	0.7016129	0.04252948	0.04191930
0.5	0.5024039	0.03315897	0.03302624
0.3	0.3025141	0.02868558	0.02687724

Nous pouvons observer que, quelle que soit la valeur des paramètres choisie, la variance empirique de $\hat{\lambda}$ est faible et on a une grande proximité entre l'écart-type théorique et l'écart-type empirique.

4.2. Problématique et résultats sur les campagnes betteravières

En France, lorsqu'un agriculteur vend des betteraves à un fabricant de sucre, la transaction se déroule en plusieurs étapes : d'abord, il arrache les betteraves du champ pour former un silo. Ensuite, des camions viennent les chercher pour les amener à l'usine. Pour chaque camion, un échantillon (d'un poids approximatif de 200 Kg) est prélevé et pesé (poids Brut 1). Cet échantillon est ensuite lavé (pour enlever la terre) et trié (pour enlever les pierres) pour être de nouveau pesé (poids Net 1). A partir de ces deux mesures, on calcule la tare terre ($100 \times$ (Net 1/Brut 1)) qui détermine le pourcentage en masse de betteraves utiles dans le camion. Par la suite, on réalise un sous-échantillon (d'un poids approximatif de 20 à 30 Kg) sur ces betteraves que l'on pèse (poids Brut 2) pour les faire décolleter par des opérateurs et obtenir le poids Net 2. Cette opération est nécessaire sur la base des accords entre le SNFS (Syndicat National des Fabricants de Sucre) et l'ARTB (Association de Recherche Technique Betteravière) compte tenu du fait qu'elle permet de déterminer quel est le pourcentage ($100 \times [1 - $(Net 2/Brut 2)]) de betteraves, appelé «taux de collet», qui ne sera pas payé par le fabricant de sucre. Par définition, le collet n'est rien d'autre que la partie supérieure de la betterave, délimitée par la dernière insertion foliaire, qui est considérée, par les 2 parties commerçantes, comme non marchande.

Le fort coût d'une telle manipulation de décolletage (qui nécessite beaucoup de personnel) aussi bien en temps qu'en argent, ainsi que la grande variance de l'opérateur ont amené le SNFS et l'ARTB à trouver une solution d'automatisation de cette mesure. Le système choisi est celui proposé par Cybernétix, une entreprise spécialiste en robotique et automatique. Le principe de mesure retenu est un rapport de volumes, estimé à partir d'une reconstitution 3D d'un vrac de betteraves par stéréovision. En pratique, le vrac de betteraves, disposé sur le tapis de réception en une couche, est photographié par un système de 3 caméras. La reconstruction 3D de la scène est obtenue par stéréo-corrélation. En aval de l'étape de reconstruction, la chaîne comprend les étapes suivantes : pré-traitement, segmentation du vrac en betteraves individuelles, analyse et tri (avec élimination des mesures «anormales») puis mesures des volumes des collets et des racines. Les traitements ne seront pas détaillés pour des raisons de confidentialité, sans toutefois que cela nuise à la compréhension de l'étude effectuée. Finalement, le taux de collet du lot obtenu est calculé par $100 \times \sum_i V_{c_i} / \sum_i (V_{c_i} + V_{r_i})$ où V_{c_i} et V_{r_i} désignent respectivement les volumes de collet et de racine calculés pour chaque betterave isolée.

Le système a permis la collecte d'une base de données. En parallèle, la mesure traditionnelle du taux de collet issue de la découpe par des opérateurs a été conservée dans les centres de réception. Le problème qui s'est posé à l'ARTB et au SNFS fut de mettre en évidence une équivalence suffisante de la mesure fournie par le système τ relativement à la mesure traditionnelle $\mathbf{t}$. Une première étude basée sur les plans d'expériences a été réalisée par Treuillet *et al.* (2004) afin d'optimiser le réglage des paramètres du système de mesure du taux de collet. Nous nous intéresserons ici à la comparaison entre la mesure système et la mesure opérateur.

L'interprétation que l'on peut donner aux différents paramètres définis dans la Section 1 est la suivante :

- $\tau_{i,j}$ est la mesure de taux de collet système du $j^{ème}$ camion du $i^{ème}$ silo.
- $\mathbf{t}_{i,j}$ est la mesure de taux de collet opérateur du $j^{ème}$ camion du $i^{ème}$ silo.
- $\mu_{s;i}$ désigne l'effet relatif à la moyenne générale, du $i^{ème}$ silo dans la valeur du taux de collet réel.
- $\mu_{g;ij}$ désigne l'effet absolu induit par le prélévement (ou camion) j dans le silo i sur le taux de collet réel, déduction faite des effets relatifs propres au silo i.
- $\varepsilon_{\tau,0;i,j}$ est l'erreur de mesure du système sur le taux de collet du $j^{ème}$ camion du $i^{ème}$ silo.
- $\varepsilon_{t,s;i}$ est l'erreur de répétabilité de l'opérateur. Celle-ci étant propre au silo lui-même.
- $\varepsilon_{t,g;i,j}$ est l'erreur de reproductibilité de l'opérateur.

Dans les Tableaux 4 et 5, nous reprenons les résultats obtenus sur les diverses campagnes. Entre 1999 et 2001, l'expérience n'a eu lieu que sur un seul site, celui d'Arcis sur Aube alors que pour l'année 2002, deux autres sites étaient expérimentés, ceux de Pithiviers et d'Origny.

TABLEAU 4
Résultats de l'étude 1999-2000-2001 à Arcis

	1999	2000	2001
I = nombre de silos	712	243	1170
$n = \sum_{i=1}^{I} n_i$ = nombre total de prélèvements	1890	1634	6323
$\widehat{\lambda}$	0.3679	0.7125	0.4836
$\widehat{\sigma}^2_{\mu,s}$ = variance de fluctuation des silos	1.72	1.09	1.59
$\widehat{\sigma}^2_{\mu,g}$ = variance des échantillons par silo	0.42	0.23	0.25
$\widehat{\sigma}^2_{\tau,0}$ = variance totale de mesure système	0.31	0.82	1.84
$\widehat{\sigma}^2_{t,g}$ = variance de reproductibilité de l'opérateur	1.72	1.86	1.76
$\widehat{\sigma}^2_{t,s}$ = variance de répétabilité de l'opérateur	0.50	1.20	1.00

TABLEAU 5
Résultats de l'étude 2002

	Arcis	Pithiviers	Origny
I = nombre de silos	130	136	80
$n = \sum_{i=1}^{I} n_i$ = nombre total de prélèvements	1038	1498	914
$\widehat{\lambda}$	0.7695	0.5490	0.5185
$\widehat{\sigma}^2_{\mu,s}$ = variance de fluctuation des silos	0.62	0.63	0.95
$\widehat{\sigma}^2_{\mu,g}$ = variance des échantillons par silo	0.19	0.48	0.24
$\widehat{\sigma}^2_{\tau,0}$ = variance totale de mesure système	1.57	1.08	0.83
$\widehat{\sigma}^2_{t,g}$ = variance de reproductibilité de l'opérateur	1.64	4.02	1.92
$\widehat{\sigma}^2_{t,s}$ = variance de répétabilité de l'opérateur	1.55	1.54	1.16

Quand on compare les résultats obtenus entre les deux tableaux, on observe une plus grande stabilité concernant les estimations de λ dans la campagne 2002. Ceci peut s'expliquer en partie par le fait que le traitement des données entre les campagnes 1999, 2000 et 2001 d'une part, et celle de 2002 d'autre part, diffère sensiblement. En effet, les trois premières années, on s'est intéressé à la normalité des couples $(\tau_i, \mathbf{t}_i)$ pour détecter d'éventuelles valeurs aberrantes, alors que, en 2002, notre intérêt s'est plutôt porté sur celle des $\mu_{i,j}$, c'est-à-dire des silos, ce qui correspond plus au modèle. De plus, le système a été légèrement modifié entre 2001 et 2002. On remarque aussi que pour le tableau 4, le rapport de I sur n est assez élevé, ce qui entraîne une certaine instabilité dans la précision des estimations.

Ces résultats sont très prometteurs et encourageants. Ils montrent que le système a de bien meilleures performances que celles qui pouvaient être prévues à partir du traitement des données globales. En effet, les valeurs de λ estimées à partir des données complètes laissaient entendre que la valeur de λ aurait pu être notoirement plus petite. Toutefois, il ressort de notre analyse que ces dernières conclusions doivent être écartées du fait que l'hétérogénéité des silos prise en compte faussait les résultats d'estimation.

5. Preuve du Théorème 1

Le but de cette section est d'établir la normalité asymptotique de l'estimateur $\widehat{\lambda}$. Pour cela nous aurons besoin des deux préliminaires suivants qui peuvent être trouvés dans Kendall *et al.*, 1987. Ces résultats étant « classiques », leur preuves seront omises.

PRÉLIMINAIRE 1. – *Si on a deux variables S_n et T_n qui sont asymptotiquement normales, et de la forme :*

$$S_n = \mu_S + \frac{\varepsilon_n}{\sqrt{n}} \quad \textit{où } \varepsilon_n \equiv \mathcal{N}(0, \sigma_1^2) \tag{84}$$

$$T_n = \mu_T + \frac{\eta_n}{\sqrt{n}} \quad \textit{où } \eta_n \equiv \mathcal{N}(0, \sigma_2^2) \tag{85}$$

on a, à l'aide d'un développement limité de S_n/T_n

$$\frac{S_n}{T_n} = \frac{\mu_S}{\mu_T} - \frac{\mu_S}{\mu_T^2}\frac{\eta_n}{\sqrt{n}} + \frac{1}{\mu_T}\frac{\varepsilon_n}{\sqrt{n}} + O_{\mathbb{P}}\left(\frac{1}{n}\right). \tag{86}$$

De plus, la variance asymptotique de S_n/T_n s'écrit :

$$\mathit{Var}\left(\frac{S_n}{T_n}\right) = \frac{\mu_S^2}{\mu_T^4}\frac{\sigma_2^2}{n} + \frac{1}{\mu_T^2}\frac{\sigma_1^2}{n} - 2\,\frac{\mu_S}{\mu_T^3}\frac{\sigma_{12}}{n} \tag{87}$$

où $\sigma_{12} = \mathit{Cov}(\varepsilon_n, \eta_n)$.

PRÉLIMINAIRE 2. – *Si* $A = (a_{ij}) \equiv W_m(n, \Sigma)$ *avec* $\Sigma = (\sigma_{ij})$ *Alors :*

$$- \mathbb{E}(a_{ij}) = n\sigma_{ij} \tag{88}$$

$$- \mathit{Cov}(a_{ij}, a_{kl}) = n\,(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) \tag{89}$$

et en notant $Vec(A) = (a_{11}, ..., a_{m1}, a_{12}, ..., a_{m2}, ...)'$, *on a :*

$$\sqrt{n}\left[Vec\Big(\frac{A}{n}\Big) - Vec\Big(\Sigma\Big)\right] \xrightarrow{d} \mathcal{N}_{m^2}(0, \Gamma) \tag{90}$$

où Γ *est la matrice de terme général*

$$\gamma_{ij,kl} = \mathit{Cov}(a_{ij}, a_{kl}) \tag{91}$$

On a en particulier, en notant $\sigma_{ii} = \sigma_i^2$ *:*

$$\mathit{Var}(a_{ii}) = 2\ n\ \sigma_{ii}^2 = 2\ n\ \sigma_i^4 \tag{92}$$

$$\mathit{Pour}\ i \neq j : \mathit{Var}(a_{ij}) = n\ (\sigma_{ii}\sigma_{jj} + \sigma_{ij}^2) = n\ (\sigma_i^2\sigma_j^2 + \sigma_{ij}^2) \tag{93}$$

$$\mathit{Cov}(a_{ii}, a_{ij}) = 2\ n\ \sigma_{ii}\sigma_{ij} = 2\ n\sigma_i^2\sigma_{ij} \tag{94}$$

Nous pouvons à présent calculer la variance asymptotique de $\widehat{\lambda}$. Il suffit d'appliquer les résultats précédents avec $S_n = T_{11} - R_{11}$ et $T_n = T_{12} - R_{12}$. Nous devons calculer d'une part les variances Var$(T_{11} - R_{11})$ et Var$(T_{12} - R_{12})$, et d'autre part la covariance Cov$(T_{11} - R_{11}; T_{12} - R_{12})$. Il est facile de noter que $(\overline{\tau}_{i,\bullet} - \overline{\tau}_{\bullet,\bullet})$ est indépendant de $(\tau_{i,j} - \overline{\tau}_{i,\bullet})$. Ceci implique que T_{11} et R_{11} sont indépendants. Un argument similaire permet de montrer que T_{12} et R_{12} sont aussi indépendants. Les deux calculs de variance découlent donc du Préliminaire 2. Toujours en utilisant l'argument d'indépendance, la covariance se réduit aux calculs de Cov$(T_{11}; T_{12})$ d'une part, et de Cov$(R_{11}; R_{12})$ d'autre part, qui découlent également du Préliminaire 2. On a donc, compte tenu des equations (36), (37), (72), (73), (88) à (94)

$$\text{Var}(T_{11}) = 2\left[\lambda^2 n_1\Big(\sigma_{\mu,s}^2 + \frac{1}{n_1}\sigma_{\mu,g}^2\Big) + \sigma_{\tau,0}^2\right]^2 \Big/ (I-1) = C/(I-1) \tag{95}$$

$$\text{Var}(R_{11}) = 2\left(\lambda^2\sigma_{\mu,g}^2 + \sigma_{\tau,0}^2\right)^2 \Big/ (n-I) = D/(I-1) \tag{96}$$

$$\begin{aligned}\text{Var}(T_{12}) = &\Bigg\{\left[\lambda^2 n_1\Big(\sigma_{\mu,s}^2 + \frac{1}{n_1}\sigma_{\mu,g}^2\Big) + \sigma_{\tau,0}^2\right]\left[n_1\sigma_{\mu,s}^2 + \sigma_{\mu,g}^2 + \sigma_{t,g}^2 + n_1\sigma_{t,s}^2\right] \\ &+ \lambda^2 n_1^2\Big[\sigma_{\mu,s}^2 + \frac{1}{n_1}\sigma_{\mu,g}^2\Big]^2\Bigg\} \Big/ (I-1) = A/(I-1)\end{aligned} \tag{97}$$

$$\begin{aligned}\text{Var}(R_{12}) =& \left\{ \left[\lambda^2\sigma^2_{\mu,g} + \sigma^2_{\tau,0}\right]\left[\sigma^2_{\mu,g} + \sigma^2_{t,g}\right] + \lambda^2\sigma^4_{\mu,g} \right\} \Big/ (n-I) \\ &= B/(I-1) \qquad (98)\end{aligned}$$

$$\begin{aligned}\text{Cov}(T_{11};T_{12}) &= 2\left\{ \lambda n_1 \left[\lambda^2 n_1 \left(\sigma^2_{\mu,s} + \frac{1}{n_1}\sigma^2_{\mu,g}\right) + \sigma^2_{\tau,0}\right]\left[\sigma^2_{\mu,s} + \frac{1}{n_1}\sigma^2_{\mu,g}\right] \right\} / (I-1) \\ &= E/(I-1) \qquad (99)\end{aligned}$$

$$\text{Cov}(R_{11};R_{12}) = 2\left\{ \lambda\sigma^2_{\mu,g}\left[\lambda^2\sigma^2_{\mu,g} + \sigma^2_{\tau,0}\right] \right\} \Big/ (n-I) = F/(I-1) \qquad (100)$$

En utilisant le Préliminaire 1, et en particulier la formule (87), et les formules (95) à (100) on déduit la variance de $\hat{\lambda}$. On en déduit la variance asymptotique donnée dans (77) car le rapport $I/(I-1)$ tend vers 1.

Remerciements

Je voudrais remercier Armelle Guillou et Pierre Cazes pour leurs commentaires pertinents.

Références

[1] DUNN O. J. & CLARK V.A. (1974), *Applied statistics : analysis of variance and regression*, John Wiley & Sons, New York-London-Sydney.

[2] EDWARDS A.L. (1979), *Multiple regression and the analysis of variance and covariance*, W. H. Freeman and Co, San Francisco.

[3] FISHER L. & MCDONALD J. (1978), *Fixed effects analysis of variance*, Academic Press, New-York, London.

[4] KENDALL M., STUART A. & ORD J.K. (1987), *Kendall's advanced theory of statistics, Vol 1, Distribution theory*, Fifth edition, The Clarendon Press, Oxford University Press, New York.

[5] MCCULLAGH P. & NELDER J.A. (1983), *Generalized linear models,* Monographs on Statistics and Applied Probability, Chapman & Hall, London.

[6] SCHEFFÉ H. (1959), *The analysis of variance*, Chapman & Hall, London.

[7] TREUILLET S., DRIOUCHI D. & RIBEREAU P. (2004), Ajustement des paramètres d'une chaîne de traitements d'images par un plan d'expériences factoriel fractionnaire 2^{k-p}, *Traitement du signal*, **21**, n°2, 141-156.